PARAMETERS TRACKING AND FAULT DIAGNOSIS BASE ON NOSQL DATABASE AT SSRF*

Y.B. Yan, Y.B. Leng[#], L.W. Lai, Z.C. Chen
Shanghai Synchrotron Radiation Facility (SSRF)
Shanghai Institute of Applied Physics, Chinese Academy of Sciences
Shanghai 201204, P.R. China

Abstract

As a user facility, the reliability and stability are very important. Besides using high-reliability hardware, the rapid fault diagnosis, data mining and predictive analytics are also effective ways to improve the efficiency of the accelerator. A beam data logging system was built at SSRF, which was based on NoSQL database. The logging system stores beam parameters under some predefined conditions. The details of the system will be reported in this paper.

OVERVIEW

Shanghai Synchrotron Radiation Facility (SSRF) is one of the advanced third generation light sources in the world, and it supports and pushes the cutting-edge scientific research and the innovation in China. It is composed of a 150 MeV linear accelerator, a 3.5 GeV booster, a 3.5 GeV storage ring, more than ten beamlines and experimental stations.

Due to the large amount of devices in the accelerator system, the era of big data is in fact coming with the rapid development of the related technology. Owing to the significant relationship among the data acquisition, storage and analysis, scale effect of big data has brought great challenges to the current data storage and analysis. Although the distributed data acquisition and processing system has been widely used, the distributed data storage and analysis has not been achieved due to the constraint of storage, network, and others. Some preliminary attempts have been making at SSRF, mainly based on the beam instrumentation and control system. This work would unearth more value of runtime data, which would also improve the stability and reliability of existing accelerator system.

SSRF beam instrumentation system consists of more than 200 devices, which covered the beam position measurement, beam charge & current measurement, beam size & length measurement, fill pattern measurement and so on [1]. All these parameters are very important during the accelerator commissioning, operation and machine studies. More than 20k scalar process variables and hundreds of 2k-points waveform records are published online per second. With proper storage and analysis tool-kits, these data could be invaluable. Otherwise the potential of various advanced electronics will be wasted.

On the other hand, various hardware and software failures have been recorded in the past few years, such as global orbit disturbance, random glitch or offset jump of individual position readings [2]. All these failures affected the reliability and stability of the accelerator system. There are no effective tools to analyze the reason due to lack of adequate raw data. The regular sampling rate of achieved data is about one hertz. History of broadband data such as turn-by-turn (several hundreds kilohertz) orbit data or bunch-by-bunch data are required in this case. Due to the huge size, the data are not likely to be stored periodically. A logging system was designed and developed, which stores the broadband data under some predefined conditions.

SYSTEM ARCHITECTURE

The data logging system is based on the Couchbase [3], which is an open source, distributed NoSQL database. It provides key-value or document access with low latency and high sustained throughput. The system architecture is shown in Figure 1.
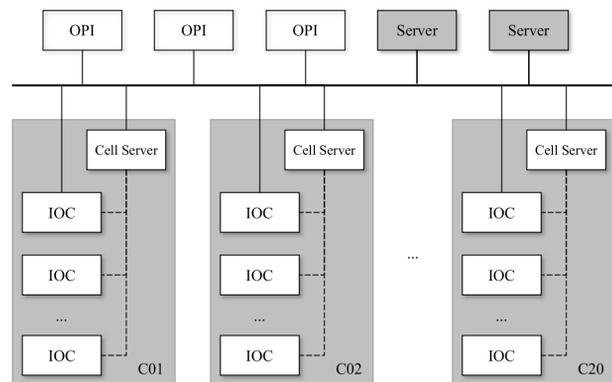

Figure 1: The architecture of data logging system.

Most part of the work is in process, especially the distributed data storage and analysis. The centralized part is mainly discussed below. In this system, IBM System x3550 M4 server and IBM Storwize V3700 storage system are adopted, which is a cost-effective option to achieve high performance. All the software run on the Linux operating system, and written using python, a widely used general-purpose, high-level programming language.

Pre-processing

For the particle accelerator, the data can be divided into two categories, hardware device related and beam related parameters. The hardware device related (such as vacuity, magnet current, undulator gap or shift, etc) and scalar beam parameters (such as beam current, close orbit, etc) have been achieved in the channel archiver. The logging system mainly stores the waveform records, including the raw data (such as turn-by-turn orbit data, bunch-by-bunch charge data, synchrotron light image, etc) and processed data (such as beam spectrum data, tune or beta function measured, etc).

A lot of data pre-processing algorithms, such as the correlation analysis, the cluster analysis and the principal component analysis, can be used to extract the useful information from the raw data. Most of the scripts are implemented using python. Beside the standard library, python supports a large number of 3rd party libraries. It makes the algorithms can be implemented easily.

Data Storage

In the current system, data are obtained under some predefined conditions, such as the global orbit disturbance, beam injection, etc. The dedicated routines are used to decide whether the conditions have been satisfied. All data under the predefined conditions are packaged and stored into the database as an entry. In order to reduce the size, the bzip2 is adopted, which is a free and open-source file compression program and uses the Burrows-Wheeler algorithm.

The distributed data storage is important for broadband data. It is the basis of high-performance parallel and distributed computing (per cell). The related research work is being carried out, which mainly focus on the Ceph [4] object store and the Gluster [5] file system.

Data Mining

The signals from various probes are different aspects of a single measurement procedure. The correlation analysis of the overall signals fits the characteristic function such as the beta function and the dispersion function to the data, which would effectively increase the usage of the original information and promote the accuracy, reliability and feasibility of the results [6].

The principal component analysis finds a small number of uncorrelated principal components that can account for the maximum amount of observed variances and covariances in the data. Each principal component is a linear combination of the observed signals and retains the maximum variance along its direction. It can be achieved by a singular value decomposition of the data matrix, such as all the turn-by-turn orbit data of the storage ring. The spatial and temporal vectors can be used to identify the betatron motion, energy motion and others, such as electronics noise [7, 8].

Result Representation

It is worth paying attention to represent the result more effectively. According to the actual demand, the report files could be automatically generated daily, or at some specific time. The matplotlib and reportlab library are adopted in python. The former is a plotting library and can save plot to image file instead of displaying it. The latter library allows rapid creation of portable document format documents. The generated files can include text (such as title, comments, calculation results, etc) and the previous generated images.

As a user facility, the reliability and stability are very important. Before the failure occurs, if the qualitative or quantitative forecasting based on the achieved data can be made, it will effectively extend the mean time between failures of the accelerator. Especially for some slow drift, the early warning will be much helpful for the operators and physicists to optimize machine parameters.

PARAMETERS TRACKING

The beam parameter tracking is very useful for the accelerator operation or machine studies, such as the tune. Now the storage ring is operated in top-up mode, which improve the efficiency and quality of synchrotron light. The top-up injections are made continuously at the time interval of about ten minutes; each injection cycle takes about ten seconds. The tune can be archived during the injections.

The tune is extracted from the excited turn-by-turn orbit data, which is the amplitude of resonance peak of betatron oscillation. The predefined condition is the gate signal of top-up injection and one entry is stored. Figure 2 shows tune drift over one hundred days. It shows that the tune is relatively concentrated in some continuous days, but scattered in different operating periods.
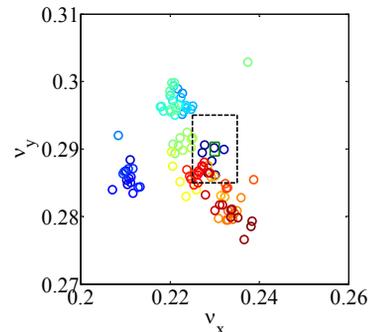

Figure 2: The tune drift during one hundred days.

The beam position monitor (BPM) system at SSRF was fully equipped with Libera Electrons. During the summer shutdown of the last few years, several units were upgraded to Libera Brilliance. Meanwhile, some other hardware and software optimization was carried out. After the upgrades, the stability and performance have been significantly improved. The track changes are shown in Figure 3.
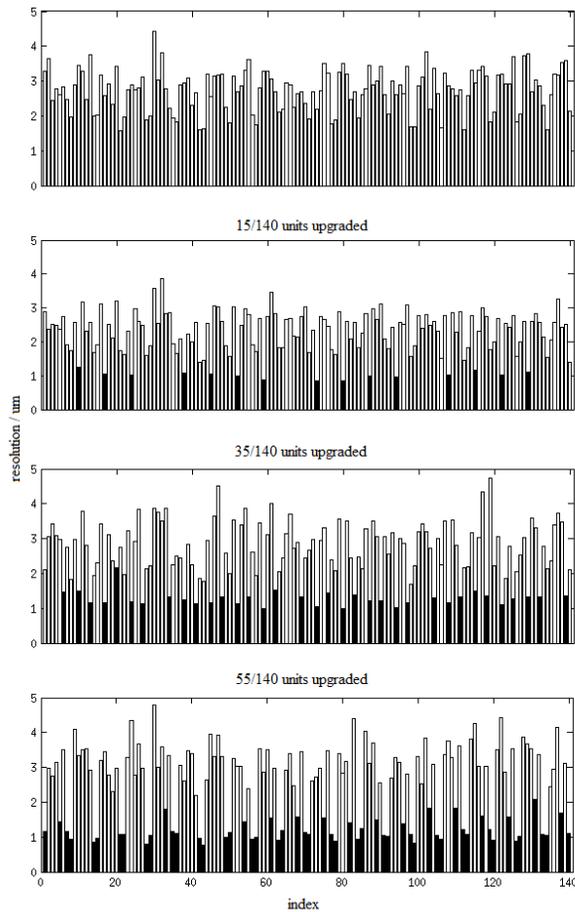
Figure 3: The track changes of electronics upgrade.

FAULT DIAGNOSIS

The particle accelerators are complicated system, with a large number of various components. The fault detection and diagnosis are a difficult task, and the data logging system will be helpful.

There is a typical example happened after a summer shutdown. The beam position monitor cables and part of the electronics were upgraded. As mentioned above, the beta function of the storage ring can be stored using some data pre-processing algorithms. But the data is abnormal at one position, shown in Figure 4. Finally the cause was found out, which is a cable connection error. The neighbouring cables (channel C and D) were cross-connected by mistake.
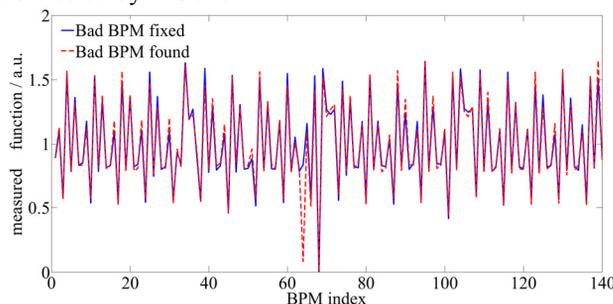

Figure 4: The beta function measured.

CONCLUSION

The beam data logging system has been designed and developed base on NoSQL database. More functionality will be added in the future. The preliminary applications show great potential for the parameters tracking and fault diagnosis. This will improve the efficiency of the operators and physicists.

With the increasing complexity of particle accelerator, the reliability and stability will become more and more crucial. Besides using high-reliability hardware, the rapid fault diagnosis, data mining and predictive analysis s are also effective ways to improve the efficiency of the accelerator.

ACKNOWLEDGMENT

REFERENCES

[1] Y. B. Leng, K.R. Ye, et al.,"SSRF Beam Diagnostics System Commissioning", DIPAC'09, Basel, Switzerland (2009).

[2] Y. B. Leng, Y. B. Yan, et al.,"Beam Diagnostics Global Data Warehouse Implementation and Application at SSRF", IPAC'11, San Sebastian, Spain (2011).

[3] Couchbase website: http://www.couchbase.com

[4] Ceph website: http://www.ceph.com

[5] GlusterFS website: http://www.gluster.org

[6] Z.C. Chen, Y. B. Leng,"Study of Correlation Analysis of Global Data from a Storage Ring", Chinese Academy of Sciences, 2014.

[7] J.H. Chen, Z.T Zhao,"Preliminary Application of Turn-by-Turn Data Analysis to the SSRF Storage Ring", Chinese Physics C, Vol. 33, No. 7, 2009.

[8] N. Zhang, Y. Yang, et al.,"Application of Model Independent Analysis Based Method to Accelerator Bunch-by-Bunch Research", High Power Laser and Particle Beams, Vol. 26, No. 3, 2014.